WR 370

Molecular Phylogenetics and Evolution
Vol. 18, No. 3, March, pp. 479–487, 2001
doi:10.1006/mpev.2000.0894, available online at http://www.idealibrary.com on IDEAL®

Evolution of Mitochondrial and Ribosomal Gene Sequences in Anophelinae (Diptera: Culicidae): Implications for Phylogeny Reconstruction

Jaroslaw Krzywinski,* Richard C. Wilkerson,† and Nora J. Besansky*

*Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556; and †Walter Reed Biosystematics Unit, Museum Support Center, Smithsonian Institution, Washington, DC

Received July 28, 2000; published online January 23, 2001

In this study, two mitochondrial genes, *cyt b* and *ND5*, and the *D2* expansion segment of the 28S nuclear ribosomal gene were used to reconstruct a phylogeny of the mosquito subfamily Anophelinae. The ingroup consisted of all three genera of Anophelinae and five of six subgenera of *Anopheles*. Six genera of Culicinae were used as the outgroup. Extreme conservation at the protein level coupled with rapid saturation of synonymous positions probably accounted for the lack of meaningful phylogenetic signal in the *cyt b* gene. In contrast, abundant variation at all codon positions of the *ND5* gene allowed recovery of the basal and most of the recent relationships. Phylogenetic analysis of *D2* produced results consistent with those of *ND5*. Combined analysis indicated well-supported monophyletic Anophelinae (with *Chagasia* basal), *Anopheles* + *Bironella*, and subgeneric clades within the genus *Anopheles*. Moreover, subgenera *Nyssorhynchus* and *Kerteszia* were supported as a monophyletic lineage. The Kishino–Hasegawa test could not reject the monophyly of *Anopheles*, whereas the recently proposed hypothesis of close affinity of *Bironella* to the subgenus *Anopheles* was rejected by the analyses of *ND5* and combined data sets. The lack of resolution of *Bironella* and *Anopheles* clades, or basal relationships among subgeneric clades within *Anopheles*, suggests their rapid diversification. Recovery of relationships consistent with morphology and previous molecular studies provides evidence of substantial phylogenetic signal in *D2* and *ND5* genes at levels of divergence from closely related species to subfamily in mosquitoes. © 2001 Academic Press

***Key Words: cyt b; ND5;* 28S rDNA; evolution; phylogeny; mosquitoes.**

INTRODUCTION

It is their vector status that puts mosquitoes (Culicidae) into a focus of particular interest to public health and to the scientific community. Among the most important diseases transmitted by those insects are dengue, filariasis, and malaria. Malaria alone, with an estimated 300–500 million clinical cases occurring annually, results in up to 3 million deaths every year. All the carriers of human malaria parasites belong to the single genus *Anopheles*, one of three genera recognized in the Culicidae subfamily Anophelinae. *Anopheles* is by far the largest, with 427 valid species and nearly worldwide distribution (Harbach, 1994). *Bironella* with 8 species is found in the Australian Region (Tenorio, 1977), and *Chagasia* with only 4 species is Neotropical (Knight and Stone, 1977). Within *Anopheles*, the smallest subgenera are Neotropical: *Kerteszia* (12 spp.), *Lophopodomyia* (6 spp.), *Nyssorhynchus* (29 spp.), and *Stethomyia* (5 spp.). The largest subgenera are *Cellia* (197 spp.), in the Old World, and *Anopheles* (178 spp.), with worldwide distribution (Harbach, 1994). Traditional approaches to the classification of Anophelinae have emphasized species discovery and taxonomic reviews of local mosquito faunas, at the expense of phylogenetic systematics (Munstermann and Conn, 1997). As a result, there lacks a reliable foundation from which to predict features important to malaria transmission or to address basic biological questions, such as sources of anopheline diversification and the influence of biogeographic forces on radiation of the group.

Only in the last few years has interest in phylogenetic studies of Culicidae grown. A phylogenetic hypothesis for the family based on morphology of fourth instar larvae, pupae, and adults was proposed by Harbach and Kitching (1998). These authors found strong evidence supporting monophyly of the subfamily Anophelinae, as well as a sister group relationship of *Anopheles* + *Bironella* and its divergence from *Chagasia*, results congruent with traditional views (e.g., Ross, 1951) and molecular phylogenies based on nuclear and mitochondrial gene sequences (Besansky and Fahey, 1997; Foley *et al.*, 1998). Furthermore, the molecular studies consistently recovered *Bironella* as basal to *Anopheles* lineages, in accord with Ross (1951). Recently, these results have been challenged by a com-

1055-7903/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2000		**00-00-2000 to 00-00-2000**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Evolution of Mitochondrial and Ribosomal Gene Sequences in Anophelinae (Diptera: Culicidae): Implications for Phylogeny Reconstruction	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Smithsonian Institution,Walter Reed Biosystematics Unit,Museum Support Center,Washington,DC,20560**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

prehensive morphological study in which *Bironella* and subgenus *Stethomyia* were recovered as a clade nested within subgenus *Anopheles* (Sallum *et al.,* 2000). Despite the extensive morphological evidence on which the study was based, this particular grouping was poorly supported. However, existing molecular evidence is also weak, due to limited taxon sampling. Only one species represented each subgenus (*Anopheles, Cellia,* and *Nyssorhynchus*) studied in Besansky and Fahey (1997), and only two subgenera, *Anopheles* and *Cellia,* the former represented by only two species, were considered in Foley *et al.* (1998). To elucidate the status of *Anopheles,* further study of the subfamily using additional data is required.

In the present study, we have explored three genes as potential markers of mosquito phylogeny: the complete sequence of the *D2* expansion segment of nuclear 28S ribosomal DNA and fragments of mitochondrial protein-coding genes cytochrome *b* (*cyt b*) and NADH dehydrogenase subunit 5 (*ND5*). Our aim was to test the utility of these genes by their ability to recover anopheline relationships strongly supported by previous studies. We also evaluated previous phylogenetic hypotheses in a statistical framework, emphasising the question of *Anopheles* monophyly.

MATERIALS AND METHODS

Taxon Sampling

The taxonomic positions and geographic sources of the 21 species analyzed are given in Table 1. The ingroup comprised 15 species (11 for the *cyt b* analysis). For four of the largest *Anopheles* subgenera, at least 2 divergent species were sampled in an effort to represent overall diversity of the group (Hillis, 1998). However, due to their rarity, only 1 or zero species from the subgenera *Stethomyia* and *Lophopodomyia,* respectively, were available. The outgroup comprised six genera of Culicinae (sensu Harbach and Kitching, 1998), among them *Aedeomyia* and *Uranotaenia,* which have been hypothesized to form a basal clade relative to all other culicine mosquitoes (Harbach and Kitching, 1998).

TABLE 1

Taxa Examined and Sources of Specimens

Anophelinae
- Genus *Anopheles* Meigen
  - Subgenus *Anopheles* Meigen
    - *quadrimaculatus* Say—Colony, Centers for Disease Control [CDC], Atlanta, USA
    - *pseudopunctipennis* Theobald—El Rama, Zelaya, Nicaragua; R. Wilkerson
    - *intermedius* (Peryassu)—Macapa, Amapa State, Brazil; R. Wilkerson
    - *mattogrossensis* Lutz and Neiva—Iquitos, Loreto, Peru; T. Klein
    - *coustani* Laveran
  - Subgenus *Cellia* Theobald
    - *stephensi* Liston—DEHLI colony, CDC
    - *gambiae* Giles—PEST colony, CDC
  - Subgenus *Nyssorhynchus* Blanchard
    - *albimanus* Wiedemann
    - *albitarsis* Lynch Arribalzaga—Baradero, Buenos Aires, Argentina; R. Wilkerson
  - Subgenus *Kerteszia* Theobald
    - *bellator* Dyar and Knab—Pariquera Acu, Sao Paulo, Brazil; R. Wilkerson
    - *cruzii* Dyar and Knab—Pariquera Acu, Sao Paulo, Brazil; R. Wilkerson
    - *neivai* Howard, Dyar and Knab
  - Subgenus *Stethomyia* Theobald
    - *kompi* Edwards—Iquitos, Peru; T. Klein
- Genus *Bironella* Theobald
  - *gracilis* Theobald—?Sepikarca, Papua New Guinea; R. Cooper
- Genus *Chagasia* Cruz
  - *bathana* (Dyar)—Belize; D. Roberts

Culicinae
- *Aedeomyia squamipennis* (Lynch Arribalzaga)—Pariquera Acu, Sao Paulo, Brazil; R. Wilkerson
- *Armigeres subalbatus* (Coquillett)—Colony, Univ. Notre Dame
- *Orthopodomyia alba* Baker—Georgia, USA; A. Cornel
- *Toxorhynchites amboinensis* (Doleschall)—Colony, Univ. Notre Dame
- *Tx. rutilus* (Coquillett)—Colony, Fla. Medical Entomol. Lab
- *Uranotaenia sapphirina* (Osten Sacken)—Georgia, USA; A. Cornel

PCR Amplification and Sequencing

Genomic DNA was extracted from individual mosquitoes (Collins *et al.,* 1987). For *cyt b,* a ~460-bp fragment representing 40% of the gene in *An. gambiae* (Beard *et al.,* 1993), was amplified with the primers cytbF (5′-GGACAAATATCATTTTGAGGAGCAACAG-3′) and cytbR (5′-ATTACTCCTCCTAGCTTATTAGGAATTG-3′) (Lyman *et al.,* 1999). For the *D2* region, primers CP12 (5′-GTGGATCCAGTCGTGTTGCTTGATAGTGCAG-3′) and CP15 (5′-GTGAATTCTTGGTCCGTGTTTCAAGACGGG-3′) were used (Porter and Collins, 1996). A ~620-bp fragment of *ND5,* representing a highly variable 3′ (37%) of the gene in *Anopheles,* was amplified from degenerate primers whose design was guided by a multiple sequence alignment of *Drosophila melanogaster, An. gambiae, An. quadrimaculatus,* and *Aedes albopictus* (Accession Nos.: M37275, L20934, L04272, and X03924, respectively). Their sequences are *ND5,* 5′-GGHYTAACTGTWWGTTATTCATTTC-3′ and tRNAPhe, 5′-CCMYAACATCTTCARTGTYAWRCTC-3′ (with IUB codes in degenerate positions). PCR amplification conditions for *D2* were described previously by Porter and Collins (1996). *Cyt b* and *ND5* were amplified in 50 μl containing 50 mM KCl, 10 mM Tris–HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 2.5 units *Taq* polymerase (GibcoBRL), 50 pmol each primer, and 1 μl template DNA (1/100th of the DNA extracted from a single mos-

TABLE 2

Number of Sites Compared for Mitochondrial *cyt b* and *ND5* and Ribosomal *D2* Sequences

	cyt b				*ND5*				*D2*	
	nt1	nt2	nt3	All sites	nt1	nt2	nt3	All sites	Total	Analyzed
Total	137	137	137	411	175	175	175	525	576	348
Variable sites	28	6	103	137	98	66	147	309	399	244
Informative sites	17	3	82	102	73	47	115	233	280	180

quito, occasionally further diluted 1:10). PCR conditions were 1–5 min initial denaturation at 94°C, followed by 35 cycles of 15–60 s at 94°C, 15–60 s at 50°C, and 45–120 s at 72°C, and a final elongation for 10 min at 72°C. On rare occasions, when amplification of *ND5* under these conditions gave poor results, hot-start PCR was performed and/or the PCR product was reamplified. One specimen per species was used for DNA amplification. The PCR products were purified with StrataPrep PCR purification kit (Stratagene) and sequenced directly with the PCR primers. Sequences of both strands were determined on an ABI 377 automated sequencer using ABI BigDye terminator chemistry (PE Applied Biosystems). Sequences have been deposited in GenBank, under Accession Nos. (*D2*) AF311226–AF311247, (*cytb*) AF311248–AF311262, and (*ND5*) AF311263–AF311281. Sequence alignments are available through NCBI (popset).

Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignment of *cyt b* and *ND5* was achieved with Clustal W 1.7 (Thompson *et al.,* 1994). Inference of character homology within *D2* was facilitated by secondary structure predictions based on the energy minimization method of Zucker (1989) as implemented by the MFold program in the GCG Sequence Analysis package (Genetics Computer Group, 1997), using the default folding temperature.

Basic sequence statistics were calculated with MEGA (Kumar *et al.,* 1993) and PAUP*4.0b2 (Swofford, 1999). Phylogenetic analyses within maximum-parsimony (MP) and maximum-likelihood (ML) frameworks were performed with PAUP*. Gaps were treated as missing characters. The minimum-parameter ML models of DNA evolution that best fit each of the data sets were determined by likelihood ratio tests using MODELTEST 2.0 (Posada and Crandall, 1998). Other than nucleotide frequencies, for which empirical values were used, ML parameters employed in the subsequent analyses were estimated iteratively from the data using PAUP*'s tree scores option, with the unweighted MP trees used for an initial estimation of the parameters. Heuristic searches were performed using 100 (ML) or 1000 (MP) random replicate addition searches with TBR branch swapping. Bootstrap support was estimated by a heuristic search on 100 (ML) or 500 bootstrap pseudoreplicates (each with 10 random additions; MP) and TBR branch swapping. To assess the extent of conflict between data sets, the incongruence length difference (ILD) test (Farris *et al.,* 1995) was implemented using PAUP* with randomizations performed 10,000 times. Alternative phylogenetic hypotheses were tested in a statistical perspective using the Kishino–Hasegawa (1989) test, in which ML trees constrained to match topologies for the hypotheses were compared to unconstrained ML trees. The hypothesis of constant evolutionary rate (molecular clock) was tested using a standard likelihood ratio test. The likelihood of the constrained tree was compared to the likelihood of the same tree with the molecular clock enforced using a χ^2 distribution with the degrees of freedom equal to the number of taxa −2 (Felsenstein, 1993).

RESULTS

Sequence Alignment and Divergence

The number of sites compared for each gene is given in Table 2. Sequences of *cyt b* were equal in length across all taxa examined, whereas one three-nucleotide deletion was observed in the *ND5* of *Tx. amboinensis.* The alignment of both protein-coding genes was unambiguous. As expected, third (nt3) and first (nt1) codon positions were more variable than nt2 positions in both genes (Table 2). However, the proportion of variable sites within the nt1 and nt2 partitions was drastically lower for *cyt b* than for *ND5.*

Substantial variability in length (340–466 bp) made the alignment of *D2* sequences problematic. Results of the automated multiple alignment using Clustal W remained controversial despite implementation of various gap opening/gap extension penalty combinations (2/1, 3/1, 4/1, 5/1, 10/1, 100/1) applied to the complete set of taxa. The final *D2* alignment roughly corresponding to the predicted secondary structure models (available on request to J.K.; also see Porter and Collins, 1996) was achieved by separate alignments of ingroup and outgroup taxa, with a gap opening/gap extension penalty of 10/1. Ingroup and outgroup alignments (561 and 418 characters long, respectively) were then aligned to each other, for a final alignment of 577

TABLE 3

Ranges of Uncorrected Pairwise Sequence Divergences (%) for *cyt b*, *ND5*, and *D2* Sequences (Completely Excluding Gaps)

Taxa	*cyt b*				*ND5*				*D2*	
	nt1	nt2	nt3	All sites	nt1	nt2	nt3	All sites	All sites	Analyzed fragments
Within Anopheles	1.5–8.1	0.0–0.0	16.1–34.3	5.8–13.4	1.7–20.1	0.0–13.8	10.3–36.2	5.1–20.8	0.8–40.3	0.3–37.4
Within Anophelinae	0.7–8.1	0.0–0.0	3.7–34.3	1.5–13.4	1.7–21.8	0.0–13.8	10.3–39.1	5.1–23.0	0.8–40.9	0.3–37.5
Within non-anophelines	2.9–8.8	0.7–2.9	18.2–33.8	8.3–14.5	13.2–26.4	8.7–19.5	24.5–37.4	15.9–27.2	12.5–26.7	10.3–23.9
Anophelinae–non-anophelines	3.7–11.0	1.5–2.2	18.9–33.6	9.0–15.4	14.9–30.5	10.9–19.0	20.1–40.8	15.7–28.0	29.2–45.4	27.5–42.6

characters. Subsequent phylogenetic analysis of the *D2* region was carried out on the 348-character data set after exclusion of the most variable fragments, in which character homology was dubious.

For all genes, there was a substantial overlap in the range of sequence divergence within Anophelinae and between anopheline taxa and other mosquitoes, suggesting a high degree of substitutional saturation (Table 3).

Phylogenetic Analysis

ND5. Because saturation plots of *ND5* suggested that the nt3 signal might be randomized with respect to phylogenetic history (data not shown), the influence of this position was examined by applying three alternative weighting schemes under a MP framework: all characters weighted equally, nt3 transitions (Ti) = 0, and nt3 = 0. Four trees were derived with equal weighting [tree length (l) = 958, consistency index excluding uninformative characters (CI) = 0.391, and rescaled consistency index (RC) = 0.168], two trees with nt3(Ti) = 0 (l = 783, CI = 0.363, RC = 0.167), and two trees with nt3 = 0 (l = 415, CI = 0.470, RC = 0.252). In general, the alternative weighting schemes affected the resolution rather than the topology of the resulting trees (Table 4).

The most appropriate ML substitution model of the *ND5* sequences chosen by MODELTEST was the general time-reversible model with among-site rate heterogeneity assumed to follow a discrete approximation of the gamma distribution (GTR + Γ; Yang, 1994). GTR allows for unequal base frequencies and unique probabilities for each of the six possible substitution classes. Under GTR, ML results were largely congruent with MP analyses. One exception was 66% support for nonmonophyly of the genus *Anopheles* in the ML bootstrap tree, in which *An. kompi* was basal to all other *Anopheles* species + *Bironella* (see Fig. 1).

D2. MP analysis resulted in a single tree (l = 647, CI = 0.544, RC = 0.327) with essentially the same topology as the ML tree derived under the HKY85 model (Hasegawa *et al.,* 1985), with among-site rate

TABLE 4

Comparative Bootstrap Support for Monophyly of Selected Groups among the Studied Taxa Supported by Alternative Data Partitions Using MP (EW, Equally Weighted; nt3(Ti) = 0, with nt3 Transitions Excluded; and nt3 = 0, nt3 Sites Excluded) and ML

Taxa	*ND5*				*D2*		*cyt b*			
	MP						MP			
	EW	nt3(Ti) = 0	nt3 = 0	ML	MP	ML	EW	nt3(Ti) = 0	nt3 = 0	ML[a]
Anophelinae	–	64	80	88	100	100	–	–	67	+
Anopheles + *Bironella*	–	65	53	72	–	–	–	–	–	–
Genus *Anopheles*	–	–	–	–	–	–	–	–	–	–
Subgenera *Anopheles* + *Cellia*	–	50	61	–	–	–	–	–	–	–
Subgenera *Kerteszia* + *Nyssorhynchus*	–	–	–	–	100	91	–	–	–	–
Subgenus *Anopheles*	–	–	–	–	81	59	–	–	–	–
Subgenus *Cellia*	89	90	91	91	99	100	–	–	–	–
Subgenus *Nyssorhynchus*	100	100	100	100	99	91	–	–	–	–
Subgenus *Kerteszia*	98	98	98	100	98	100	–	–	–	–
An. cruzii + *An. bellator*	87	71	–	–	–	–	86	86	59	+

[a] Bootstrap support not calculated; + denotes presence of the lineage in the tree; – denotes lack of the lineage.

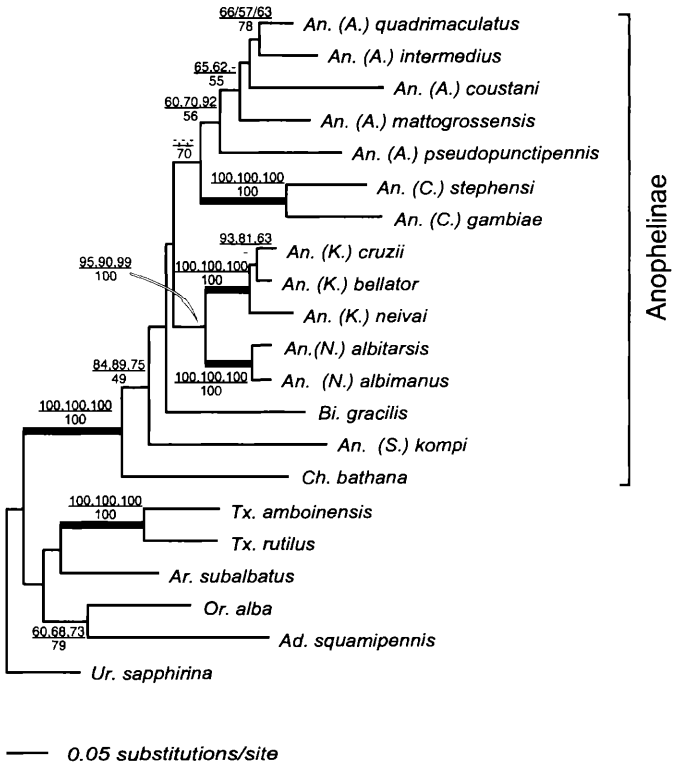

FIG. 1. Phylogenetic relationships within mosquitoes based on combined *D2* and *ND5* data sets. The ML tree (−ln L = 8751.45) was inferred using the GTR model with different rates of substitution for each substitution class (A ↔ G = 0.7987, C ↔ T = 4.411, A ↔ C = 3.232, A ↔ T = 3.046, C ↔ G = 2.879, G ↔ T = 1), among-site rate variation accommodated by assuming a proportion of sites to be invariable (0.1739), and rates at the remaining sites assumed to follow a gamma distribution (α = 1.0332). Numbers at nodes represent bootstrap support (>50%) for the ML analysis (below line) and for three weighting schemes of *ND5* data (equal weighting, nt3Ti = 0, and nt3 = 0) for the combined MP analysis (above line). Thick lines represent branches with 100% bootstrap support in all four analyses.

variation assumed to follow the gamma distribution (HKY85 + Γ). This model allows for unequal base frequencies and different substitution rates of transitions and transversions. The *D2* data resolved some of the relationships within *Anopheles,* which remained unresolved on the *ND5* bootstrap trees: subgenus *Anopheles* was recovered for the first time, and subgenera *Nyssorhynchus* and *Kerteszia* were strongly supported as sister taxa (Table 4).

Cyt b. One most-parsimonious tree was produced from a MP analysis of all positions weighted equally (l = 326, CI = 0.405, RC = 0.176), 13 with nt3(Ti) = 0 (l = 193, CI = 0.368, RC = 0.178), and 16 with nt3 = 0 (l = 44, CI = 0.523, RC = 0.354). No substantial differences resulted from ML analysis with the GTR model, with among-site rate variation accommodated by assuming some proportion of sites to be invariable, and rates at the remaining sites assumed to follow the gamma distribution (GTR + I + Γ; Gu *et al.,* 1995; Table 4). Very poor recovery of morphology-based groups and conflict with the largely congruent *D2* and *ND5* trees suggests that homoplasy in the *cyt b* data has obscured the phylogenetic signal, a suggestion supported by saturation plots (not shown). For these reasons we excluded *cyt b* from the combined phylogenetic analysis, as its main contribution would be noise.

Combined data. The ILD test (Farris *et al.,* 1995) showed that *ND5* and *D2* data sets were combinable, although the significance level was close to rejection of the null hypothesis (P = 0.055). MP analysis of the combined data under all three weighting schemes recovered monophyletic Anophelinae, *Anopheles* + *Bironella, Nyssorhynchus* + *Kerteszia,* and four subgenera of *Anopheles: Anopheles, Cellia, Kerteszia,* and *Nyssorhynchus* (Fig. 1).

The ML analysis under a GTR + I + Γ model yielded a topology different from that of the MP trees with regard to the positions of *Bironella* and subgenus *Stethomyia* relative to other subgeneric clades of *Anopheles.* However, MP and ML bootstrap analyses gave largely congruent results. Two exceptions were the lack of ML bootstrap support for the *Anopheles* + *Bironella* clade and the lack of MP support for subgenera *Celia* + *Anopheles* as sister taxa (Fig. 1).

Hypothesis testing. The present analyses did not recover traditional relationships within Anophelinae, where *Anopheles* is regarded a monophyletic taxon, nor those proposed by Sallum *et al.* (2000) in which *Bironella* is nested within subgenus *Anopheles.* According to the Kishino–Hasegawa test, the traditional phylogeny, with *Bironella* as sister taxon to *Anopheles,* was not in significant conflict with the most likely trees derived from our data sets (Table 5). Similarly, the alternative hypothesis (*Bironella* within subgenus

TABLE 5

Results of Kishino–Hasegawa Test for Alternative Hypotheses (Two-Tailed Test at the 5% Significance Level)

	Genus *Anopheles* monophyletic			*Bironella* within subgenus *Anopheles*		
Data set	Diff. —ln L	SD	*P*	Diff. —ln L	SD	*P*
ND5	3.513	4.021	0.3827	31.121	13.232	0.0190
D2	6.254	5.386	0.2463	18.053	9.190	0.0503
ND5 + D2	1.579	4.809	0.7428	36.281	11.604	0.0018

Anopheles) was not significantly different from the unconstrained *D2* ML tree, although the *P* value (0.0503) was very close to the level of rejection. In contrast, *ND5* and combined data rejected the position of *Bironella* among the species of subgenus *Anopheles*.

DISCUSSION

Dynamics of Sequence Divergence and Phylogenetic Utility

ND5 is one of the fastest-evolving mitochondrial genes (Clary and Wolstenholme, 1985) and for this reason its phylogenetic signal is likely to be degraded by multiple changes when more diverged taxa are compared. Parsimony may be an inconsistent estimator of relationships under such conditions due to inability of the method to detect multiple substitutions on long branches. Performance of parsimony, however, can be improved by implementation of more realistic evolutionary models via differential weighting based on the information about potential saturation drawn from the data. Indeed, after elimination of likely saturated sites as judged by saturation plots (nt3 transitions or the entire nt3 partition; data not shown), congruence with previous studies (Besansky and Fahey, 1997; Harbach and Kitching, 1998; Sallum *et al.,* 2000) regarding basal clades increased, albeit at the expense of decreased resolution at most shallow divergences.

Substantial phylogenetic information was retrieved from *D2* despite the large sequence divergence observed among more distantly related taxa and the short fragment studied. The present analysis was based on conserved sequence fragments, most of which were predicted to be involved in maintaining core stem structure (Michot and Bachellerie, 1987; Porter and Collins, 1996; Ruiz Linares *et al.,* 1991). Although *D2* provided very high support for Anophelinae and most relationships within *Anopheles,* this data set did not recover well-supported relationships among *Anopheles, Bironella,* and *Chagasia.* Further refinement of the alignment based on secondary structure might help resolve these relationships (Kjer, 1995). Alternatively, the basal anopheline lineages might have rapidly diversified from each other, consistent with very short basal branches of the *D2* ML tree (data not shown). This possibility is incompatible with the results from the *ND5* ML analysis and combined parsimony analysis for the divergence of *Chagasia,* which was joined to other Anophelinae by a relatively long and well-supported branch. A nonmutually exclusive explanation is that *D2* may not have evolved in a clock-like manner, but rather at a very slow rate during the early history of Anophelinae and then at an accelerated rate within several individual ingroup lineages. In support of this, the likelihood ratio test strongly rejected the assumption of a molecular clock ($P \ll 0.01$), a result consistent with highly unequal branches in the *D2* tree. Increased evolutionary rates and/or different ages of extant lineages studied may explain high observed divergence within Anophelinae relative to the outgroup (Table 3).

A striking feature of *D2* is the conservation of its sequences in closely related species of the subgenera *Kerteszia* and *Nyssorhynchus* compared to *ND5.* For example, there were 32 differences between two species in *Nyssorhynchus* observed in *ND5* and only 5 in the complete *D2* sequences. Similarly, the number of differences between *An. cruzii* and *An. bellator,* closely related species belonging to the subgenus *Kerteszia,* was threefold higher in *ND5* relative to *D2.* A practical implication of low variability in *D2* is lack of phylogenetic signal strong enough to recover the latter species in a single clade. Strong differences in the number of changes and in the lengths of corresponding branches on ML trees (data not shown) indicate that *D2* evolves at a slower rate than *ND5* in those lineages. Whether the rate in *D2* is slower than that of other genes in mosquitoes is not yet known. There is evidence that this is the case in *Drosophila,* in which sibling species in the melanogaster group have identical *D2* sequences (Ruiz Linares *et al.,* 1991) but are clearly distinguishable using IGS and ETS regions of the rDNA unit (Tautz *et al.,* 1987) or protein-coding genes, such as alcohol dehydrogenase (Bodmer and Ashburner, 1984).

Most of the relationships recovered with *cyt b* were at odds with the accepted taxonomy, earlier hypotheses of mosquito phylogeny, or the results from the other genes employed in this study. Apparently, this is due to lack of phylogenetic information in that gene resulting from extreme conservation of its protein product, dictated by strong functional constraints (Meyer, 1994). Whereas synonymous changes became fully saturated, not enough nonsynonymous changes have accumulated (34 variable and 20 parsimony informative of 274 sites; Table 2) to provide meaningful phylogenetic signal. Similarly, studies of relationships in insect families Braconidae (Hymenoptera) (Gimeno *et al.,* 1997) and Staphylinidae (Coleoptera) (Ballard *et al.,* 1998) showed no or little phylogenetic utility of *cyt b* gene fragments. In kissing bugs (Hemiptera: Reduviidae: Triatominae), however, analysis of the same *cyt b* region as studied here produced highly supported relationships consistent with the accepted taxonomy (Lyman *et al.,* 1999).

Phylogenetic Relationships

In agreement with the previous studies, Anophelinae was strongly supported by our data as a monophyletic group. The basal split within the subfamily into *Chagasia* and *Anopheles* + *Bironella* clades was also moderately supported by the *ND5* and the combined analysis. *Chagasia* has been traditionally considered an ancient taxon of anophelines, as it shows several characters reminiscent of nonanopheline mosquitoes,

such as strongly arched mesonotum, trilobed scutellum, and presence of setae on the postpronotum (Belkin, 1962).

Anopheles was not separated from *Bironella* as a single clade using present data, which reflects close relationships between both genera. As Belkin (1962) stated, *Bironella* "is very similar to *Anopheles* in the immature stages and differs in the adults only in minor palpal, venational and genitalic characters." Earlier molecular studies based on the *white* (Besansky and Fahey, 1997) and mitochondrial *COII* (Foley *et al.,* 1998) genes recovered monophyly of *Anopheles* relative to *Bironella.* Both studies employed only a limited number of *Anopheles* subgenera, which could have resulted in easier recovery of the genus. In the case of *COII,* strong support for monophyletic *Anopheles* was achieved only after successive approximation weighting. All these studies probably do not represent a genuine conflict about the placement of *Bironella,* as our topologies were not significantly different from monophyly of *Anopheles,* the traditional hypothesis (Table 5). The difference between the best tree and the tree with (*Bironella* + *Stethomyia* + subgenus *Anopheles*) constraint was also not significant for the *D2* analysis. In contrast, *ND5* and the combined data rejected the hypothesis of Sallum *et al.* (2000) that *Bironella* and *Stethomyia* are nested within the subgenus *Anopheles.* None of the trees inferred from the present data recovered this grouping. Instead, subgenus *Anopheles* was recovered either as monophyletic using the *D2* and combined data or in association with *Cellia* using the *ND5* data. The position of *Stethomyia* and *Bironella* nested within subgenus *Anopheles* is poorly supported and based on homoplasious morphological characters (Sallum *et al.,* 2000); thus, it may not reliably represent true relationships within the group. Further molecular evidence based on new *white* gene data (Krzywinski *et al.,* 2001) also contradicts the *"Bironella* within *Anopheles"* hypothesis and gives additional support for the monophyly of *Anopheles.* The apparent conflict between traditional phylogeny and our present results most likely reflects insufficient resolving power of the data to give conclusive order to the branches in question (Kishino and Hasegawa, 1989).

This study failed to recover basal relationships within *Anopheles,* despite the recovery of highly supported subgeneric clades in most cases. Foley *et al.* (1998) suggested that subgenus *Anopheles* is not monophyletic, because in their study *An. quadrimaculatus,* a member of this subgenus, was placed among *Cellia* species. However, their observation was poorly supported even after successive approximation weighting and was based on highly A + T-biased (92.8%) nt3 positions that were likely saturated. Cunningham (1997) showed that the successive approximation approach is strongly influenced by the starting tree, and the method reinforces its topology whether the tree is correct or not. Moreover, the placement of *An. quadrimaculatus* may be an artifact of a too-distant outgroup, *Drosophila yakuba.*

A close relationship between subgenera *Anopheles* and *Cellia* was suggested by the *white* gene study of Besansky and Fahey (1997) and morphological studies of Sallum *et al.* (2000), a suggestion supported by our results. Weighted parsimony analysis of the *ND5* gene clustered subgenera *Anopheles* and *Cellia* together and the combined ML analysis recovered both lineages as sister taxa with moderate bootstrap support. In addition, the *D2* data and combined analyses revealed close affinities between subgenera *Nyssorhynchus* and *Kerteszia.* This relationship was hypothesized by Root (1922), and according to some authors (e.g., Edwards, 1932) *Kerteszia* was even included in the subgenus *Nyssorhynchus.* The taxonomic status of *Kerteszia* was not clear to Zavortink (1973), who was the last reviewer of this subgenus, but Peyton *et al.* (1992) found numerous diagnostic characters for both taxa, consistent with their present status as independent subgenera. Finally, Sallum *et al.* (2000) recovered strongly supported *Nyssorhynchus* and *Kerteszia.* Very high bootstrap values for both clades in our study also support their present subgeneric status.

CONCLUSIONS

The present study provides evidence of substantial phylogenetic signal in *D2* and, to a lesser extent, in *ND5* sequences at levels of divergence from closely related species to subfamily in mosquitoes. Both genes provided results largely congruent with each other and with the combined data; likewise, different methods of analysis resulted in similar topologies. The areas of conflict between the trees were limited to branches that, in general, received very poor support from either data set. Good resolution was achieved when analysis of the combined data sets was conducted. The study revealed, however, several undesirable properties of the data, which can confound phylogenetic reconstruction of more distant relationships or within some ingroup lineages. Those were strong base composition bias, saturation at many sites, and unequal nucleotide frequencies across taxa in *ND5* and unequal rates of substitution among lineages together with alignment difficulties resulting from a specific mode of evolution in *D2.* Adding more species to the analysis may reduce some of the mentioned problems, e.g., with the alignment or with unequal rates in *D2.* A denser sampling in the lineages with long branches may break them up and lead to a higher resolution of the basal relationships within *Anopheles* + *Bironella.* Despite some negative properties, both genes from the sample of taxa studied gave results consistent not only with each other and with the new evidence from the *white* gene (Krzywinski *et al.,* 2001) but also largely consistent

with earlier studies, proving their usefulness as phylogenetic markers in Anophelinae. *Cyt b,* on the other hand, appears to lack phylogenetic signal for the relationships other than those among very closely related species.

The phylogenetic analysis indicated well-supported monophyletic Anophelinae (with *Chagasia* occupying basal position), *Anopheles* + *Bironella,* and subgeneric clades within the genus *Anopheles.* Moreover, subgenera *Nyssorhynchus* and *Kerteszia* were supported as a monophyletic lineage. The Kishino–Hasegawa test could not reject the monophyly of *Anopheles,* whereas the recently proposed hypothesis of *Bironella* nested within the subgenus *Anopheles* was rejected by the analyses of *ND5* and combined data sets. The lack of resolution of *Bironella* and *Anopheles* clades, as well as basal relationships among subgeneric clades within *Anopheles,* suggests their rapid diversification. However, to approach a better picture of anopheline evolution, additional information from independent loci is needed. Protein-coding single-copy nuclear genes are promising candidates.

ACKNOWLEDGMENTS

We gratefully acknowledge R. Cooper, A. Cornell, and D. Roberts for providing valuable specimens from their collections. We thank Y. Ling for technical support. K. McCracken, R. DeSalle, and several anonymous reviewers (!) made valuable comments on an earlier version of the manuscript. N.B. received support from the UNDP/World Bank/WHO Special Programme and Training in Tropical Diseases, from the John D. and Catherine T. MacArthur Foundation, from the National Institutes of Health, and from the University of Notre Dame.

REFERENCES

Ballard, J. W. O., Thayer, M. K., Newton, A. F., and Grimser, E. R. (1998). Data sets, partitions, and characters: Philosophies and procedures for analyzing multiple data sets. *Syst. Biol.* **47:** 367–396.

Beard, C. B., Hamm, D. M., and Collins, F. H. (1993). The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* **2:** 103–124.

Belkin, J. N. (1962). "The Mosquitoes of the South Pacific (Diptera: Culicidae)," Vol. 1. Univ. of California Press, Berkeley.

Besansky, N. J., and Fahey, G. T. (1997). Utility of the *white* gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). *Mol. Biol. Evol.* **14:** 442–454.

Bodmer, M., and Ashburner, M. (1984). Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila. Nature* **309:** 425–430.

Clary, D. O., and Wolstenholme, D. R. (1985). The mitochondrial DNA molecule of *Drosophila yacuba:* Nucleotide sequence, gene organization and genetic code. *J. Mol. Evol.* **22:** 252–271.

Collins, F. H., Mendez, M. A., Rasmussen, M. O., Mehaffey, P. C., Besansky, N. J., and Finnerty, V. (1987). A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.* **37:** 37–41.

Cunningham, C. W. (1997). Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Syst. Biol.* **46:** 464–478.

Edwards, F. W. (1932). Diptera fam. Culicidae. *In* "Genera Insectorum" (P. Wytsman, Ed.). Desmet-Verteneuil, Brussels.

Farris, J. S., Källersjö, M., Kluge, A. G., and Bult, C. (1995). Testing significance of incongruence. *Cladistics* **10:** 315–319.

Felsenstein, J. (1993). PHYLIP (Phylogeny inference package), version 3.5. Department of Genetics, University of Washington, Seattle.

Foley, D. H., Bryan, J. H., Yeates, D., and Saul, A. (1998). Evolution and systematics of *Anopheles:* Insights from a molecular phylogeny of Australasian mosquitoes. *Mol. Phylogenet. Evol.* **9:** 262–275.

Genetics Computer Group (1997). Program manual for the GCG package. Version 9.1. Madison, WI.

Gimeno, C., Belshaw, R., and Quicke, D. L. J. (1997). Phylogenetic relationships of the Alysinae/Opiinae (Hymenoptera: Braconidae) and the utility of *cytochrome b,* 16S and 28S *D2* rRNA. *Insect Mol. Biol.* **6:** 273–284.

Gu, X., Fu, Y., and Li, W. (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* **12:** 546–557.

Harbach, R. E. (1994). Review of the internal classification of the genus *Anopheles* (Diptera: Culicidae): The foundation for comparative systematics and phylogenetic research. *Bull. Entomol. Res.* **84:** 331–342.

Harbach, R. E., and Kitching, I. J. (1998). Phylogeny and classification of the Culicidae (Diptera). *Syst. Entomol.* **23:** 327–370.

Hasegawa, M., Kishino, M., and Yano, T. (1985). Dating the human–ape split by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Hillis, D. M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47:** 3–8.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29:** 170–179.

Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* **4:** 314–330.

Knight, K. L., and Stone, A. (1977). "A Catalog of the Mosquitoes of the World (Diptera: Culicidae)," 2nd ed., Vol. 6, Thomas Say Found. College Park, MD.

Krzywinski, J., Wilkerson, R. C., and Besansky, N. J. (2001). Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: Insights from nuclear single copy genes and the weight of evidence. *Syst. Biol.,* in press.

Kumar, S., Tamura, K., and Nei, M. (1993). MEGA: Molecular evolutionary genetics analysis, version 1.01. Pennsylvania State Univ., University Park.

Lyman, D. F., Monteiro, F. A., Escalante, A. A., Cordon-Rosales, C., Wesson, D. M., Dujardin, J.-P., and Beard, C. B. (1999). Mitochondrial DNA sequence variation among triatomine vectors of Chagas' disease. *Am. J. Trop. Med. Hyg.* **60:** 377–386.

Meyer, A. (1994). Shortcomings of the *cytochrome b* gene as a molecular marker. *Trends Ecol. Evol.* **9:** 278–280.

Michot, B., and Bachellerie, J. P. (1987). Comparisons of large subunit rRNAs reveal some eukaryote-specific elements of secondary structure. *Biochimie* **69:** 11–23.

Munstermann, L. E., and Conn, J. E. (1997). Systematics of mosquito disease vectors (Diptera, Culicidae): Impact of molecular biology and cladistic analysis. *Annu. Rev. Entomol.* **42:** 351–369.

Peyton, E. L., Wilkerson, R. C., and Harbach, R. E. (1992). Compar-

ative analysis of the subgenera *Kerteszia* and *Nyssorhynchus* of *Anopheles* (Diptera: Culicidae). *Mosquito Syst.* **24:** 51–69.

Porter, C. H., and Collins, F. H. (1996). Phylogeny of Nearctic members of the *Anopheles maculipennis* species group derived from the *D2* variable region of 28S ribosomal RNA. *Mol. Phylogenet. Evol.* **6:** 178–188.

Posada, D., and Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Root, F. M. (1922). The classification of American *Anopheles* mosquitoes. *Am. J. Hyg.* **2:** 321–322.

Ross, H. H. (1951). Conflict with *Culex. Mosquito News* **11:** 128–132.

Ruiz Linares, A., Hancock, J. M., and Dover, G. A. (1991). Secondary structure constraints on the evolution of *Drosophila* 28S ribosomal RNA expansion segments. *J. Mol. Biol.* **219:** 381–390.

Sallum, M. A. M., Schultz, T. R., and Wilkerson, R. C. (2000). Phylogeny of Anophelinae based on morphological characters. *Ann. Entomol. Soc. Am.* **93:** 745–775.

Swofford, D. L. (1999). PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4.0b2. Sinauer, Sunderland, MA.

Tautz, D., Tautz, C., Webb, D. A., and Dover, G. A. (1987). Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: Implications for molecular coevolution in multigene families. *J. Mol. Biol.* **195:** 525–542.

Tenorio, J. A. (1977). Revision of the genus *Bironella* (Diptera: Culicidae). *J. Med. Entomol.* **14:** 317–361.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **11:** 316–324.

Zavortink, T. J. (1973). Mosquito studies (Diptera, Culicidae) XXIX: A review of the subgenus *Kerteszia* of *Anopheles. Contrib. Am. Entomol. Inst.* **9:** 1–54.

Zucker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244:** 48–52.